Correlation between faecal microbial taxa and ulcerative colitis in different phases of disease activity in a north Indian cohort

**Garima Juyal^{1*#}, Ajit Sood², Vandana Midha³, Arshdeep Singh², Dharmatma Singh²,
Ramit Mahajan², Vijay Verma^{4\$}, Rakesh Bhatnagar^{1§}, Mohan C Joshi⁵**

Affiliations

1. School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

2. Department of Gastroenterology, Dayanand Medical College and Hospital,
Ludhiana, India

3. Department of Medicine, Dayanand Medical College and Hospital, Ludhiana, India

4. Division of Medical Bio-informatics, Indian Council of Medical Research, New
Delhi 110029, India

5. Multidisciplinary Centre for Advanced Research and Studies (MCARS), Jamia
Millia Islamia, New Delhi, India.

[#]Present Address- Department of Genetics, University of Delhi South Campus,
Benito Juarez Marg, New Delhi, India

^{\$}Present Address- North Florida Research and Education Centre (NFREC)/Institute
of Food and Agricultural Sciences (UF/IFAS), University of Florida, Quincy 32351,
FL, USA

[§]Present Address-Amity University Rajasthan, Jaipur

***Corresponding author**

21

22

23 Abstract

24 **Background and Aims:** A link between gut microbiota and Ulcerative colitis (UC) has been
 25 established in several studies. However, only a few studies have examined specific changes
 26 in microbiota associated with different phases of disease activity in UC. In this study, we
 27 attempted to identify differentially abundant taxa between (i) UC patients and healthy
 28 controls and (ii) different phases of disease activity, in a genetically distinct north Indian (NI)
 29 population.

30 **Methods:** 16S rRNA (V3–V4) sequencing of 105 patients with UC [newly diagnosed
 31 (n=14); patients in remission (n=36) and active disease (relapse, n=55)]; and 36 healthy
 32 controls was performed and analysed.

33 **Results:** Lower gut microbial diversity; enrichment of lactate-producing bacteria namely
 34 *Streptococcus*, *Bifidobacterium* and *Lactobacillus*; and depletion of butyrate-producing
 35 bacteria (e.g., *Lachnospiraceae* and *Ruminococcaceae*), were observed among UC patients.
 36 Subgroup analysis revealed differential abundance of *Escherichia-Shigella*, *Streptococcus*,
 37 *Enterococcus* and *Faecalibacterium* in newly diagnosed UC patients. No discrete microbial
 38 features were observed between patients in remission and those with active disease. Co-
 39 occurrence network analysis revealed a mutualistic association between opportunistic
 40 pathogens and *Bifidobacterium* and *Lactobacillus* and antagonistic relationship with butyrate-
 41 producers.

42 **Conclusion:** This first faecal microbiome study among NI UC cohort elucidated
 43 dysanaerobiosis; loss of short chain fatty acid producers and enrichment of inflammation
 44 associated microbes; population specific differential microbial genera; and a potential
 45 microbial signature for early dysbiosis.

46

47

48 **Introduction**

49 Historically considered to be a disease of the western world, ulcerative colitis (UC), has
 50 emerged as a global disease ¹. The etiology of UC is multifactorial. However it is believed
 51 that a complex dynamic interplay between genetic factors, environmental influences and
 52 intestinal microbiota drives chronic inflammation, characteristic of UC ^{2,3}. The conventional
 53 treatment strategies focus on inducing and maintaining remission and preventing disease
 54 related complications by targeting the dysregulated immune system ⁴. In the past decade
 55 however, the intestinal microbiota has unfolded as a crucial integrant in the pathogenesis of
 56 UC. Novel microbiota targeted therapies therefore have made inroads into the current
 57 treatment paradigms as alternative/adjunctive therapies ^{5,6}.

58 Several microbiome studies, primarily from Western countries, have demonstrated differences
 59 in the faecal microbial profile between UC patients and healthy individuals ⁷. However there
 60 is paucity of data from developing countries. In light of the differences in genetic
 61 predisposition to IBD across ethnically divergent populations ⁸, and their interactions with
 62 non-genetic factors predominantly diet, on the development and clinical spectrum of UC, it is
 63 imperative to survey the gut microbiome from genetically, culturally and socially divergent
 64 populations.

65 In the current study, we analysed the faecal microbial composition in UC patients and healthy
 66 controls of north Indian (NI) origin. Since UC is an emerging disease in developing countries,
 67 in contrast to the western world where it has reached a plateau (prevalence equilibrium), we
 68 propose that characterizing the gut microbiome of a UC cohort from a developing country like
 69 India provides a unique opportunity to identify early microbial markers.

70

71 **Materials and Methods**

72 **Cohort**

73 Consecutive patients (age >18 years) with an established diagnosis of UC⁹ attending the
 74 outpatient clinic at Dayanand Medical College and Hospital, Ludhiana, India between
 75 January 2016 and December 2017 were categorized into two groups i.e. **newly diagnosed**
 76 [patients with active disease (defined as Mayo clinic score >2) diagnosed within four weeks
 77 of enrolment with no prior exposure to treatment for UC (including corticosteroids, 5-
 78 aminosalicylates (5-ASA), immunomodulators and biologics)]; and **previously diagnosed**
 79 (diagnosed >6 months prior to enrolment and on standard of care pharmacotherapy for UC).
 80 Patients in this group were further divided into two sub-groups: **relapse:** previously
 81 diagnosed UC with active disease and **remission:** previously diagnosed UC in clinical
 82 remission (defined as Mayo clinic score ≤ 2 , with each sub-score ≤ 1 and endoscopic sub-
 83 score of 0). In patients with clinical symptoms of active disease, a limited sigmoidoscopy was
 84 performed for endoscopic evidence of disease activity. Infections like *Clostridioides difficile*,
 85 cytomegalovirus (CMV) and/or Epstein Barr virus (EBV) were tested in patients with active
 86 disease. Clinical details including patient demographics, disease characteristics and treatment
 87 history were recorded. Patients with age <18 years, Crohn's colitis, indeterminate colitis, *C.*
 88 *difficile* infection, co-infection with CMV or EBV, co-morbid illnesses such as severe heart,
 89 lung, or neurological disease, and active malignancies, antibiotic use during the month before
 90 inclusion and diagnosis of UC between 4 weeks to 6 months prior to enrolment were
 91 excluded. Adult (age >18 years), unrelated volunteers who had no comorbidities or disorders
 92 known to be associated with changes in gut microbiota, served as **healthy controls**.

93

94 **Faecal DNA extraction and 16S rRNA sequencing**

95 Faecal samples were collected from patients and healthy controls at the hospital settings. The
96 samples were labelled using unique patient specific codes, homogenized and stored at -80°C
97 within 2 hours. DNA was extracted from faecal samples as a non-invasive proxy for the gut
98 microbiome using DNeasy PowerLyzer PowerSoil Kit (**Qiagen/MO BIO cat# 12855-50**) as
99 per manufacturer's protocol. For profiling microbiome composition, the V3-V4 hyper-
100 variable region of the bacterial 16S rRNA gene was sequenced on Illumina Miseq platform
101 using a commercial facility.

102 **Data analysis**

103 Paired-end reads were merged using FLASH¹⁰ with a minimum and maximum overlap of
104 30bp and 250bp respectively. Low base quality (average $Q \geq 34$) reads were removed using
105 fastx toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). De-replication was
106 performed in VSEARCH¹¹ for the identification of unique sequences. A *de novo* approach in
107 VSEARCH¹¹ was used to remove chimeric sequences. Bacterial community composition of a
108 sample was achieved using closed_reference Operational Taxonomic Unit (OTU) picking
109 method in UCLUST¹². In this method sequencing reads were clustered based on 97%
110 similarity to form OTU and subsequently one representative sequence from each OTU was
111 picked. Any OTU that had a count of 1 in a single sample (identified as singletons) was
112 removed. Taxonomy assignment was done using RDP classifier¹³ against Silva-132 database
113 at 97% similarity. OTU table was then normalized to the same depth. Alpha and Beta
114 diversity indices were estimated and distance metrics were represented as principal
115 coordinate analysis (PCoA) plots using QIIME1.9.1¹⁴. Hierarchical clustering and Co-
116 occurrence network analysis were performed using ClustVis¹⁵ and SParse InversE
117 Covariance Estimation for Ecological Association Inference (SPIEC-EASI) program in R¹⁶
118 respectively. Both these methods are detailed in **Supplementary text**.

119 Shapiro–Wilk normality test was used to test the distribution of data. The bacterial numbers
120 were transformed into relative percentages for statistical analysis. Significant differences in
121 alpha and beta diversity were tested using Mann-Whitney T-test and PERMONOVA
122 respectively. Significant differences in the relative abundance between groups were
123 computed using Mann-Whitney test and $p < 0.05$ was considered statistically significant after
124 Bonferroni correction.

125

Results

Patient characteristics and 16S rRNA amplicon sequencing data

A total of 105 UC patients comprising three subgroups i.e. newly diagnosed, n=14; Relapse, n=55; Remission, n=36 and 36 healthy controls were included in the study (**Table 1**). Sequencing of V3-V4 region of faecal DNA samples yielded a total of 114200323 ± 478237.89 (sum $\pm$ SD) paired-end reads with a read length of 268.37 ± 23.80 (mean $\pm$ SD) [**Table 2a**]. After removing low quality reads, chimera sequences and singleton, a total of 300467 ± 221183.15 (mean $\pm$ SD) and 458981.08 ± 112054.93 (mean $\pm$ SD) sequencing reads were obtained in UC patients and healthy controls, respectively [**Table 2a**]. The sequence reads were normally distributed across samples ($p < 0.001$) [**Table 2b**].

Microbial composition in UC patients differs from that of controls

Sequencing reads were rarefied to 20,000 sequences per sample to control for variations in sequencing depth. Significant low alpha diversity was observed in UC patients compared to healthy controls [Shannon Index ($P_{\text{corrected}} < 9.7E-14$); Chao1 ($P_{\text{corrected}} < 3.5E-14$); number of observed species ($P_{\text{corrected}} < 3.5E-14$) [**Figures. 1a-c**]. Beta diversity showed significant dissimilarity in the compositional structure of bacterial communities between UC patients and healthy controls (**PERMANOVA** $p = 0.001$; **Figure 2**). Tight clustering in healthy controls indicated high gut microbial compositional similarity among themselves compared to UC patients [**Figure 2a**]. However, a scattered distribution among UC represented heterogeneity underlying disease pathogenesis [**Figure 2b**]. No differences in diversity was observed among the three UC subgroups. However, significant alpha diversity differences were observed between the three UC subgroups and healthy controls [**Supplementary Figs. 1-2**].

Overall distribution of predominant phyla that contributed to >99% of the total taxonomy based on median abundance in UC patients and healthy controls is depicted in **Figure 3a**. The most abundant phylum in healthy controls was Firmicutes which accounted for $58.4 \pm 8.1\%$ (median abundance $\pm$ SD) followed by Bacteroidetes ($21.8 \pm 9.6\%$), Actinobacteria ($9.1 \pm 10.9\%$) and Proteobacteria ($2.1 \pm 4.1\%$). On the other hand in UC cohort, Firmicutes contributed $57.7 \pm 19.9\%$ followed by Actinobacteria ($18.3 \pm 14.0\%$), Bacteroidetes ($5.14 \pm 20.4\%$) and Proteobacteria ($1.7 \pm 11.4\%$) [**Figure 3a**]. Of these, the relative abundance of Actinobacteria, Bacteroidetes and Firmicutes significantly differed ($P_{\text{corrected}} < 0.05$) between UC patients and healthy controls [**Figure 3b**]. When compared independently with healthy controls, Bacteroidetes and Actinobacteria were found significantly altered across the three UC subgroups. In addition, significant difference in the abundance of Bacteroidetes ($P_{\text{corrected}} = 0.05$) was observed between newly diagnosed and remission groups (**Figs. 4a-b**).

At the genus level, we observed 11 genera in UC patients and 20 genera in healthy controls representing >1.0 % of the total OTUs (**Figure 5**) which reiterates higher diversity and richness in the healthy group. Among these, relative abundance of 18 were significantly altered [$P_{\text{corrected}} < 0.05$] in UC patients compared to healthy controls [**Figure 6a**]. In the subgroup analysis, we observed nine genera in newly diagnosed, 13 in remission and 11 in relapse which represented >1.0 % of the total OTUs [**Figure 5b**]. Of these, only four namely uncultured_Bacteroidetes, *Escherichia-Shigella*, *Enterococcus*, and *Ruminococcaceae* were significantly altered [$P_{\text{corrected}} < 0.05$] across the three subgroups [**Figure 6b**]. Furthermore, on comparing the three UC subgroups independently with healthy controls, 19 genera were found to be significantly altered ($P_{\text{corrected}} < 0.05$) (**Figure 6b**). The p-value of altered taxa and their abundance are presented in **Table 3**.

To validate the differential abundance of nitrate reducing bacteria in the faeces of UC patients and healthy controls nitrate reduction activity was determined using nitrate broth [**detailed in**

supplementary text]. Significant increase in the number of nitrate reducing bacteria per gram of faeces was observed in UC patients compared to healthy controls ($P_{\text{corrected}}=E-08$). However, no differences were observed across the three UC subgroups [**Supplementary Figure 3**].

Hierarchical clustering reveals newly diagnosed UC as a unique clad

Clustering analysis categorized UC patients and healthy controls into distinct clads based upon differential OTU abundance (>1%) at genus level and heat maps are represented in **Fig 7a**. Barring four, the remaining genera were found to be statistically significant in the abundance analysis [**Table 3**]. Further, cluster analysis also illustrated that newly diagnosed UC patients formed a distinct clad compared to remission, relapse and healthy controls [**Fig 7b**].

Co-occurrence network analysis reveals enrichment of pathobionts in UC patients

A total of 115-116 pruned OTUs were included for co-occurrence network analysis to study their potential relationships in different pathological contexts among UC patients. Such analysis gives an initial cue on the structure of the community and how metabolite cross-feeding/metabolic dependency and nutritional preferences contribute to microbial assemblage and their maintenance. This analysis revealed (a) *Streptococcus*, *Escherichia-Shigella*, *Bifidobacterium*, *Lactobacillus* and *Faecalibacterium* in newly diagnosed; (b) *Streptococcus*, *Lactobacillus*, *Prevotella* in relapse; (c) *Streptococcus*, *Blautia* and *Erysipelotrichaceae* in remission; and d) *Ruminococaccae*, *Bifidobacterium*, *Prevotella* and *Bacteroides* in healthy controls, to be the predominant taxa in the networks [**Figure 8**].

Discussion

Gut microbial dysbiosis in terms of alteration in composition, reduction in biodiversity and consequently loss of metabolic and immune homeostasis has been associated with development of UC¹⁷. To date, most of the published studies on UC from India have been largely based on either quantification of specific bacteria in tissue biopsy or faecal samples¹⁸⁻²⁰ or 16S rRNA analysis of limited tissue biopsy samples^{19,21}. To our knowledge, this report is the first description of faecal microbiota profile of NI UC patients which provides a snapshot of the alterations of the microbiota in different phases of disease.

We report that the ethnically distinct NI UC population exhibited a marked reduction in the bacterial diversity and significant alteration in the structure of faecal bacterial community as compared to healthy controls [**Figures. 1 & 2**] lending support to the well noted central feature of dysbiosis in UC pathology. We found significant loss of genera comprised of dominant anaerobes alongside expansion of subdominant facultative anaerobes suggestive of a disruption in anaerobiosis in the gut lending support to oxygen hypothesis²², discussed in the later section. Each of these genera has been discussed below in detail. Findings from subgroup analyses have been discussed in the subsequent section.

UC patients vs healthy controls

Loss of SCFA producing taxa: We observed a significant loss of anaerobic SCFA producing bacteria in UC patients compared to healthy controls [**Table 3; Figure 6a**]. This is in consonance with previous studies from European and Asian populations^{7,23,24}. SCFAs are organic products mainly composed of acetate, propionate, and butyrate produced after fermentation of dietary fibers and resistant starches in the colon. They act as signaling molecules and regulate different biological processes, including the promotion of gut integrity, immune response, reduction of pathogenic bacterial population; and prevention of unwanted infiltration of bacteria from lumen to lamina propria^{25,26}. This suggests that lower

abundance of SCFA-producing bacteria in UC patients may favour a shift towards an inflammation-promoting microbiome, thereby enhancing host inflammation and modifying the disease behaviour. This is in line with oxygen hypothesis which posits that chronic inflammation of intestinal walls results in extravasation of blood from ulcerated tissue which leads to increased release of hemoglobin carrying oxygen and reactive oxygen species into the intestinal lumen, which in turn creates a microenvironment that favors expansion of facultative anaerobes. The resulting decrease in anaerobes (which release anti-inflammatory compounds) leads to increased inflammation, establishing a positive feedback loop that accelerates the disease process ²².

Enrichment of potentially harmful bacteria: A significant enrichment of *Streptococcus*, a lactic acid-producing bacteria (LAB) was observed in UC patients [Table 3, Figure 6] which is in line with earlier reports ²⁷⁻²⁹. Some species of this bacterium have shown to induce or promote pro-inflammatory response in macrophages, human tissue, and intestine ²⁹ suggesting that metabolites and structural components of *Streptococcus* and immune cells are at direct/indirect interplay and lead to excessive inflammation. In a recent study, increased abundance of f_Streptococcaceae among IBD patients was correlated with proton pump inhibitors (PPIs) ^{3,30}. However, in our case none of the UC patients were using PPIs suggesting other factors responsible for *Streptococcus* abundance. Interestingly, evidences are now emerging and pointing to association between higher levels of gut inflammation in IBD patients and increase in bacteria typical of the oral cavity ²⁸. Although oral microbes are innately resistant to colonization in healthy guts, it has been proposed that inflammation or strain-specific adaptation (including antibiotic resistance and virulence genes) of oral microbes allows their colonization in the gut resulting in exaggerated inflammatory response ³¹. Therefore, it would be interesting to further explore the link between oral microbiome and UC.

246 The other enteric anaerobes which were found to be enriched in UC patients were also LABs
 247 namely *Bifidobacterium* and *Lactobacillus* [Table 3, Figure 6]. These findings are in stark
 248 contrast to majority of earlier studies from west ^{7,23,32}. However, enrichment of these bacteria
 249 have been reported in the faecal and tissue biopsy samples of active UC patients from
 250 Chinese and Russian populations ^{33,34}. Such observed differences reinforce the potent role of
 251 host genetics in determining the composition of the gut microbiota. This derives support from
 252 recent reports documenting a) that differences in microbiota are highly impacted by the ethnic
 253 background ³⁵; b) microbiota is partly heritable as evidenced in studies analysing both
 254 concordant and discordant twins for IBD ³⁶; c) influence of genomic loci on several microbial
 255 genera in mice ³⁷; and d) specific variants of the *NOD2* associated with changes in the
 256 abundance of the f_Enterobacteriaceae in IBD patients ³⁸. Importantly, evidences now
 257 emerging pinpoint that these commensals have the potential to become opportunistic or
 258 promote colonization of other pathobionts and elicit immunostimulatory effects ^{39,40}. For
 259 instance, a recent study in mice demonstrated that the introduction of a single commensal
 260 species, *Lactobacillus reuteri*, was sufficient to exacerbate experimental autoimmune
 261 encephalomyelitis in a genetically susceptible host ⁴¹. This suggests that these “good”
 262 commensals can become “bad” and can cause unwarranted immune stimulation in genetically
 263 susceptible individuals reiterating the likely effect of host genetics on microbiome
 264 composition. Other than host genetics, differential selection pressure through diet ^{42,43};
 265 disease characteristics; and persistent use of 5-ASA (which changes the colonic luminal pH
 266 and hence promote growth of *Bifidobacterium* and *Lactobacillus*) ³⁵, could be the non-genetic
 267 factors that may have contributed on the gut microbial assemblage. Considering the global
 268 uniformity of treatment modalities it seems that effect of disease modifying drugs on
 269 expansion of these genera is minimal. A few recent studies which have shown that disease

activity, 5-ASA dosage or addition of glucocorticoids had no impact on the microbial stability
⁴⁴ lend support to this.

Subgroup analysis

Early dysbiosis hallmark

To avoid the confounding effects of long-standing disease, disease modifying agents (such as medication), and co-morbid conditions which could have potential effect on microbiome composition, treatment naïve newly diagnosed UC patients were analysed in relation to treatment experienced patients [remission and relapse] and healthy controls in this study. An over-representation of *Escherichia-Shigella*, *Enterococcus* and *Streptococcus* in newly diagnosed UC patients which is in line with prior studies from both adult ⁷ and paediatric ⁴⁵ UC patients was notable [Table 3]. *Escherichia-Shigella* and *Enterococcus* have been reported to have the ability to adhere to and invade intestinal epithelial cells, leading to an inflammatory immune response ⁴⁶. Besides other SCFA producing taxa, *Faecalibacterium*, a dominant species in the gut microbiota of healthy subjects and which has anti-inflammatory properties was found to be significantly depleted in newly diagnosed UC patients [Table 3]. The reduction of extremely oxygen-sensitive *Faecalibacterium* and increase in facultative anaerobes reiterates increased oxidative stress in lumen (oxygen hypothesis, discussed above). *F. prausnitzii*, the sole species of *Faecalibacterium* has been suggested to be an important regulator of intestinal homeostasis ⁴⁷ and its association with IBD has been established in several gut microbial studies ^{7,23}.

Similar to studies investigating paediatric patients with IBD, the association of *Enterococcus*, *Escherichia*, *Streptococcus*, and *Faecalibacterium* with newly diagnosed adult UC patients observed in this study may represent an early dysbiotic shift triggering disease. This may imply that these genera could be considered as early non-invasive faecal biomarkers for

294 epithelial dysfunction which includes inflammation, aggravated intestinal injury and
295 increased intestinal permeability.

296

No microbial shift between UC patients in remission and relapse

Gut microbial profiling during different phases of disease activity in UC has been explored in only a few studies with inconsistent and conflicting results⁷²³. In the present study, no significant differences were found between disease activity and faecal microbiota (except uncultured_Bacteroidetes) [Table 3; Figures 6b, 7b]. Consistent with our observation, a few earlier studies have shown no microbial differences between inactive and active UC patients^{44,48}. Additionally, no significant alteration in the bacterial abundances in relation to disease activity were observed in UC patients of European origin³. Similar trends were also observed in two other independent longitudinal studies wherein gut microbiota in UC patients remained highly stable regardless of disease stage, disease activity or treatment escalation^{44,49}. Thus, associations identified in previous cross-sectional studies may more likely reflect inter-individual variation rather than disease activity. No discrete microbial features between remission and relapse groups may highlight the high level of host genetic and gut microbial heterogeneity. Therefore, to unravel robust shifts in the microbiota related to disease activity species/strain mapping and metabolite profiling may be essential.

***Streptococcus*- a persistent pathobiont across the three UC subgroups**

No differential abundance of *Streptococcus* was observed across the three UC subgroups. However, on comparing these subgroups independently with healthy controls we observed significant enrichment of *Streptococcus* in newly diagnosed patients and those in relapse [Table 3]. Although not statistically significant, *Streptococcus* was found to be higher in remission group compared to healthy [Figure 6]. The enrichment of *Streptococcus* across the three UC subgroups [Figs. 6-8] suggest that colonization of *Streptococcus* is not transient but resilient (despite treatment regimen or dietary modifications) and thus has a continued role in disease pathology. Of note, the positive correlation of *Streptococcus* with *Lactobacillus*, *Bifidobacterium* and *Escherichia-Shigella* as evident in our co-occurrence analysis [Figure 8]

suggests mutualistic association among them and validates the relative abundance findings [Figure 6; Table 3].

Enrichment of nitrate reducing bacteria as an indicator of inflammation

Culture-based analysis showed an increased abundance of nitrate reducing bacteria in the faeces of UC patients compared to healthy controls. These bacteria are capable of converting nitrate to nitrite which are further reduced to nitric oxide and ammonia (nitrogen reactive species). It is now widely believed that oxidants, including free radicals, such as nitric oxide, play a key role in the initiation and perpetuation of inflammation and in the subsequent tissue damage in IBD^{50,51}. No significant differences in nitrate reducing activity was observed among the three UC subgroups, however the activity in each subgroup was found to be significantly higher compared to healthy controls [Table 3; Supplementary Figure 3]. While *Escherichia-Shigella* and *Enterococcus* attributed to a higher nitrate reducing activity in newly diagnosed UC patients, *Bifidobacterium* and *Lactobacillus*, which have been recently reported to have nitrate reducing activity⁵², may explain higher activity among patients in remission and relapse. Interestingly, the latter derives support from a recent study from India wherein significant enrichment of *Lactobacilli* was observed in the faecal samples of active UC patients compared to healthy controls and their level significantly reached close to controls during remission^{19,32}. These observations confirm increased oxidative stress in the gut lumen; and independently validates our observations.

Acknowledging the modest sample size, the significant associations of bacterial taxa (as detailed above) with UC, identified in the present study require replication in larger, independent cohorts of matched ethnicity. However, the high concordance between our findings and several other reports discussed above clearly indicate that our study is well powered. Furthermore, the limitation that 16S rRNA sequence analysis does not enable species/strain assignment, high-resolution mapping of gut microbiome in NI UC patients is

highly encouraged. This will unravel more robust disease endotype specific associations driven by select strains which may serve as biomarker. The search for relevant uncultured/unnamed species will expand our knowledge on their relevance in the pathogenesis of UC.

In summary, this first sequencing based faecal microbiome analysis in a NI UC cohort revealed reduced bacterial diversity and altered composition. Most of the genera identified in our study have shown recurring association with UC across ethnically and geographically distinct populations despite differences in methodology, bio-specimens and reference databases used. Our observations may imply that a) restoration of health promoting SCFA producing bacteria could be a pragmatic solution for mitigating disease; and b) *Escherichia-Shigella*, *Enterococcus* and *Streptococcus* could serve as non-invasive microbial signatures for inflammation of high clinical utility. Higher abundance of *Bifidobacterium* and *Lactobacillus* among NI UC patients suggest inherent population specific differences in disease pathogenesis driven by host genetics, diet and environmental attributes. Their relationship with UC pathogenesis however warrants further investigations. A deeper understanding of how bacterial genera identified in this study are involved in interactions with host genetics, diet and lifestyle factors etc. either by regulating microbial pathways or by production of specific metabolites may facilitate identification of robust prognostic and diagnostic markers and for better microbial based therapeutic interventions.

370 **Ethical consideration**

371 The Institutional ethics review board of Jawaharlal Nehru University, New Delhi and
372 Dayanand Medical College and Hospital, Ludhiana approved the study (IEC numbers
373 2017/SERB Young Scientist/125 and DMCH/R&D/2015/238 respectively) approved the
374 study. Written informed consent was obtained from each participant.

375 **Funding**

376 The financial assistance for this work was provided by Science and Engineering
377 Research Board, New Delhi, vide F. NO. SB/YS/LS-191/2014 to GJ.

378 **Acknowledgements**

379 The authors thank Prof. Thelma BK, Department of Genetics, University of Delhi
380 South Campus for critical discussions throughout the study; MedGenome Labs Ltd.
381 for 16S rRNA sequencing service; Mr. Vikas for logistic support at Dayanand
382 Medical College and Hospital, Ludhiana, India; MCJ acknowledges funding support
383 from DBT/Wellcome Trust Intermediate grant (IA/I/15/2/502086).

384 **Data availability**

385 The data underlying this article will be shared on reasonable request to the corresponding
386 author.

387 **Competing interests**

388 The authors declare that they have no competing interests.

389 **Authors' contributions**

390 GJ, AjS and VM conceived and designed the study; AjS and VM recruited clinical
391 samples and clinical data; GJ performed faecal DNA isolation, data analysis,
392 interpretation of data, wrote the first draft and secured funding; AS and RM
393 contributed to patient follow-up, clinical data acquisition; DS to sample collection

394 and clinical data acquisition; VV set up the data analysis pipeline and performed
395 independent data analysis; RKB provided infrastructure and other support for
396 conducting the study to GJ as a SERB fellow; MCJ provided critical scientific and
397 technical inputs throughout the study; All authors contributed to manuscript writing
398 and approved the final submitted version.
399

Legends

Table 1. Baseline clinical characteristics of UC patients enrolled in the study

Table 2 Sequencing reads obtained a) before and after quality control and b)

distribution of sequencing reads after quality control

Table 3. Relative abundance of top genera (>1% prevalence) in healthy controls,

total UC patients and UC subgroups

Figure. 1 Alpha diversity metrics of fecal bacterial communities in UC and

healthy controls. Whisker box plots illustrate (A) species diversity; and (B) species

richness between the UC patients and healthy controls. Rarefaction curve (C) shows

the observed species at various sequencing depths in UC patients and healthy

controls. * represents Bonferroni corrected Mann-Whitney p value.

Figure. 2 Beta diversity analysis based on the overall structure of the faecal

microbiota of UC patients and healthy controls.

Principal Coordinate Analysis (PCoA) of UC patients and healthy controls based on

(A) unweighted and (B) weighted UniFrac metric and (C) Bray Curtis dissimilarity.

Percentages on the x-axis and y-axis correspond to the percent variation in gut

microbiome compositions explained by PC1 and PC2. The gut microbiota of UC

patients show significant dissimilarity from healthy controls (p=0.001

PERMONOVA).

Figure. 3 Taxonomic diversity of faecal microbiota composition at the phylum

level among UC patients compared to healthy (A) Distribution of faecal

microbiota composition and (B) Inter-individual variation of predominant phyla

based on their relative abundance. * represents Bonferroni corrected Mann-Whitney p

value.

Figure. 4 Taxonomic diversity of faecal microbiota composition at the phylum level within UC subgroups compared to healthy (A) Distribution of faecal microbiota composition and (B) Inter-individual variation of predominant phyla based on their relative abundance. * represents Bonferroni corrected Mann-Whitney p value.

Figure. 5 Taxonomic diversity of faecal microbiota composition at the genus level (A) total UC patients and (B) UC subgroups compared to healthy * represents Bonferroni corrected Mann-Whitney p value.

Figure. 6 Differential abundances of microbial genera representing >1.0% of total OTUs (A) UC patients and (B) UC subgroups compared to healthy. * represents Bonferroni corrected Mann-Whitney p value.

Figure. 7 Hierarchical clustering analysis of genera representing >1.0% of total OTUs (A) all UC patients and (B) UC subgroups compared to healthy. The color code indicates relative abundance of OTUs ranging from red (high) to blue (low). The OTUs were transformed to $(\ln(x + 1))$ values, rows were centered and vector scaling was applied using ClustVis.

Figure. 8 Co-occurrence networks analysis of microbial genera based upon its relative abundance (A) Newly Diagnosed, (B) Relapse, (C) Remission and (D) Healthy. Each node represents genus and is uniquely colored. The size of each node is proportional to the abundance of the genus in log scale. Line between two nodes represents correlation; green and red lines indicate positive and negative association respectively. Predominant network in UC subtypes and healthy controls is represented with dotted circle.

Supplementary Figure. 1 Alpha diversity metrics of fecal bacterial communities in UC subgroups and healthy controls. Whisker box plots represent comparison of

(A) species diversity and (B) species richness between the three UC subgroups and healthy controls. Rarefaction curve (C) shows the observed species at various sequencing depths in three UC subgroups and healthy controls. * represents Bonferroni corrected Mann-Whitney p value.

Supplementary Figure. 2 Beta diversity analysis based on the overall structure of the fecal microbiota between UC subgroups and healthy controls

Principal Coordinate Analysis (PCoA) of UC subgroups i.e. newly diagnosed (red), remission (blue) and relapse (orange) and healthy controls (green) based on (A) unweighted and (B) weighted UniFrac metric and (C) Bray Curtis dissimilarity. Each point in the plots represents the composition of the faecal microbiome of one individual and distances indicate degree of similarity to other individuals. The closer the spatial distance of the sample, the more similar the species composition of the sample. Percentages on the x-axis and y-axis correspond to the percent variation in gut microbiome compositions explained by PC1 and PC2. No significant differences among the three UC subgroups were observed.

Supplementary Figure. 3 Relative nitrate reducing activity

Whisker box plots illustrate relative absorbance (620nm) per gram of faeces observed in nitrate broth of (A) UC patients and healthy and (B) UC subgroups. Each dot represents relative absorbance in a sample. The bottom, middle, and top boundaries of each box represent the first, second (median), and third quartiles of the relative absorbance. The whiskers (lines extending from the top and bottom of the box and ending in horizontal cap) extend to points within 1.5 times the interquartile range. The points extending above the whiskers are outliers. *represents Mann-Whitney Bonferroni corrected p value.

References

1. Snell A., Segal J., Limdi J., Banerjee R. Inflammatory bowel disease in India: challenges and opportunities. *Frontline Gastroenterol* 2020;flgastro-2020-101500. Doi: 10.1136/FLGASTRO-2020-101500.
2. Loddo I., Romano C. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front Immunol* 2015;6. Doi: 10.3389/fimmu.2015.00551.
3. F I., A VV., MJ B., J F., D G., MC V., et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 2018;67(1):108–19. Doi: 10.1136/GUTJNL-2016-312135.
4. Dubinsky MC. Reviewing treatments and outcomes in the evolving landscape of ulcerative colitis. *Postgrad Med* 2017;129(5):538–53. Doi: 10.1080/00325481.2017.1319730.
5. de Fátima Caldeira L., Borba HH., Tonin FS., Wiens A., Fernandez-Llimos F., Pontarolo R. Fecal microbiota transplantation in inflammatory bowel disease patients: A systematic review and meta-analysis. *PLoS One* 2020;15(9). Doi: 10.1371/JOURNAL.PONE.0238910.
6. Ghouri YA., Richards DM., Rahimi EF., Krill JT., Jelinek KA., DuPont AW. Systematic review of randomized controlled trials of probiotics, prebiotics, and synbiotics in inflammatory bowel disease. *Clin Exp Gastroenterol* 2014;7:473–87. Doi: 10.2147/CEG.S27530.
7. Pittayanon R., Lau JT., Leontiadis GI., Tse F., Yuan Y., Surette M., et al. Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel Diseases: A Systematic Review. *Gastroenterology* 2020;158(4):930-946.e1. Doi: 10.1053/J.GASTRO.2019.11.294.

- 498 8. Park SC., Jeon YT. Genetic Studies of Inflammatory Bowel Disease-Focusing on
499 Asian Patients. *Cells* 2019;**8**(5):404. Doi: 10.3390/cells8050404.
- 500 9. Maaser C., Sturm A., Vavricka SR., Kucharzik T., Fiorino G., Annese V., et al.
501 ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis,
502 monitoring of known IBD, detection of complications. *J Crohns Colitis*
503 2019;**13**(2):144–64. Doi: 10.1093/ECCO-JCC/JJY113.
- 504 10. Magoč T., Salzberg SL. FLASH: Fast length adjustment of short reads to improve
505 genome assemblies. *Bioinformatics* 2011;**27**(21):2957–63. Doi:
506 10.1093/bioinformatics/btr507.
- 507 11. Rognes T., Flouri T., Nichols B., Quince C., Mahé F. VSEARCH: A versatile open
508 source tool for metagenomics. *PeerJ* 2016;**2016**(10). Doi: 10.7717/peerj.2584.
- 509 12. Edgar RC. Search and clustering orders of magnitude faster than BLAST.
510 *Bioinformatics* 2010;**26**(19):2460–1. Doi: 10.1093/bioinformatics/btq461.
- 511 13. Wang Q., Garrity GM., Tiedje JM., Cole JR. Naïve Bayesian classifier for rapid
512 assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ*
513 *Microbiol* 2007;**73**(16):5261–7. Doi: 10.1128/AEM.00062-07.
- 514 14. Caporaso JG., Kuczynski J., Stombaugh J., Bittinger K., Bushman FD., Costello EK.,
515 et al. QIIME allows analysis of high-throughput community sequencing data. *Nat*
516 *Methods* 2010:335–6. Doi: 10.1038/nmeth.f.303.
- 517 15. Metsalu T., Vilo J. ClustVis: A web tool for visualizing clustering of multivariate data
518 using Principal Component Analysis and heatmap. *Nucleic Acids Res*
519 2015;**43**(W1):W566–70. Doi: 10.1093/nar/gkv468.
- 520 16. Kurtz ZD., Müller CL., Miraldi ER., Littman DR., Blaser MJ., Bonneau RA. Sparse
521 and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS*

- 522 *Comput Biol* 2015;**11**(5). Doi: 10.1371/journal.pcbi.1004226.
- 523 17. Nishida A., Inoue R., Inatomi O., Bamba S., Naito Y., Andoh A. Gut microbiota in the
524 pathogenesis of inflammatory bowel disease. *Clin J Gastroenterol* 2018. Doi:
525 10.1007/s12328-017-0813-5.
- 526 18. Verma R., Verma AK., Ahuja V., Paul J. Real-time analysis of mucosal flora in
527 patients with inflammatory bowel disease in India. *J Clin Microbiol*
528 2010;**48**(11):4279–82. Doi: 10.1128/JCM.01360-10.
- 529 19. Kedia S., Rampal R., Paul J., Ahuja V. Gut microbiome diversity in acute infective
530 and chronic inflammatory gastrointestinal diseases in North India. *J Gastroenterol*
531 2016:660–71. Doi: 10.1007/s00535-016-1193-1.
- 532 20. Kumari R., Ahuja V., Paul J. Fluctuations in butyrate-producing bacteria in ulcerative
533 colitis patients of North India. *World J Gastroenterol* 2013;**19**(22):3404–14. Doi:
534 10.3748/wjg.v19.i22.3404.
- 535 21. Walujkar SA., Kumbhare S V., Marathe NP., Patangia D V., Lawate PS., Bharadwaj
536 RS., et al. Molecular profiling of mucosal tissue associated microbiota in patients
537 manifesting acute exacerbations and remission stage of ulcerative colitis. *World J*
538 *Microbiol Biotechnol* 2018;**34**(6). Doi: 10.1007/s11274-018-2449-0.
- 539 22. Henson MA., Phalak P. Microbiota dysbiosis in inflammatory bowel diseases: In silico
540 investigation of the oxygen hypothesis. *BMC Syst Biol* 2017;**11**(1):1–15. Doi:
541 10.1186/S12918-017-0522-1/FIGURES/6.
- 542 23. Caenepeel C., Tabib NSS., Vieira-Silva S., Vermeire S. Review article: how the
543 intestinal microbiota may reflect disease activity and influence therapeutic outcome in
544 inflammatory bowel disease. *Aliment Pharmacol Ther* 2020;**52**(9):1453–68. Doi:
545 10.1111/APT.16096.

- 546 24. Zakerska-Banaszak O., Tomczak H., Gabryel M., Baturo A., Wolko L., Michalak M.,
547 et al. Dysbiosis of gut microbiota in Polish patients with ulcerative colitis: a pilot
548 study. *Sci Rep* 2021;**11**(1):2166. Doi: 10.1038/s41598-021-81628-3.
- 549 25. Venegas DP., De La Fuente MK., Landskron G., González MJ., Quera R., Dijkstra G.,
550 et al. Short chain fatty acids (SCFAs) mediated gut epithelial and immune regulation
551 and its relevance for inflammatory bowel diseases. *Front Immunol*
552 2019;**10**(MAR):277. Doi: 10.3389/FIMMU.2019.00277/BIBTEX.
- 553 26. Deleu S., Machiels K., Raes J., Verbeke K., Vermeire S. Short chain fatty acids and its
554 producing organisms: An overlooked therapy for IBD? *EBioMedicine* 2021;**66**. Doi:
555 10.1016/J.EBIOM.2021.103293.
- 556 27. Alam MT., Amos GCA., Murphy ARJ., Murch S., Wellington EMH., Arasaradnam
557 RP. Microbial imbalance in inflammatory bowel disease patients at different
558 taxonomic levels. *Gut Pathog* 2020 *121* 2020;**12**(1):1–8. Doi: 10.1186/S13099-019-
559 0341-6.
- 560 28. Bartlett A., Gullickson RG., Singh R., Ro S., Omaye ST. The Link between Oral and
561 Gut Microbiota in Inflammatory Bowel Disease and a Synopsis of Potential Salivary
562 Biomarkers. *Appl Sci* 2020, *Vol 10, Page 6421* 2020;**10**(18):6421. Doi:
563 10.3390/APP10186421.
- 564 29. Heidarian F., Noormohammadi Z., Asadzadeh Aghdaei H., Alebouyeh M. Relative
565 abundance of Streptococcus spp. and its association with disease activity in
566 inflammatory bowel disease patients compared with controls. *Arch Clin Infect Dis*
567 2017;**12**(2). Doi: 10.5812/archcid.57291.
- 568 30. Imhann F., Bonder MJ., Vila AV., Fu J., Mujagic Z., Vork L., et al. Proton pump
569 inhibitors affect the gut microbiome. *Gut* 2016;**65**(5):740–8. Doi: 10.1136/GUTJNL-

- 570 2015-310376.
- 571 31. Schirmer M., Denson L., Vlamakis H., Franzosa EA., Thomas S., Gotman NM., et al.
572 Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative
573 Colitis Patients Are Linked to Disease Course. *Cell Host Microbe* 2018;**24**(4):600-
574 610.e4. Doi: 10.1016/J.CHOM.2018.09.009.
- 575 32. Kedia S., Ahuja V. Epidemiology of Inflammatory Bowel Disease in India: The Great
576 Shift East. *Inflamm Intest Dis* 2017;**2**(2):102–15. Doi: 10.1159/000465522.
- 577 33. Gryaznova MV., Solodskikh SA., Panevina AV., Syromyatnikov MY., Dvoretzkaya
578 YD., Sviridova TN., et al. Study of microbiome changes in patients with ulcerative
579 colitis in the Central European part of Russia. *Heliyon* 2021;**7**(3):e06432. Doi:
580 10.1016/j.heliyon.2021.e06432.
- 581 34. Wang W., Chen L., Zhou R., Wang X., Song L., Huang S., et al. Increased proportions
582 of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria
583 in inflammatory bowel disease. *J Clin Microbiol* 2014;**52**(2):398–406. Doi:
584 10.1128/JCM.01500-13.
- 585 35. Ventura RE., Iizumi T., Battaglia T., Liu M., Perez-Perez GI., Herbert J., et al. Gut
586 microbiome of treatment-naïve MS patients of different ethnicities early in disease
587 course. *Sci Reports* 2019 91 2019;**9**(1):1–10. Doi: 10.1038/s41598-019-52894-z.
- 588 36. Willing BP., Dicksved J., Halfvarson J., Andersson AF., Lucio M., Zheng Z., et al. A
589 pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with
590 inflammatory bowel disease phenotypes. *Gastroenterology* 2010;**139**(6). Doi:
591 10.1053/J.GASTRO.2010.08.049.
- 592 37. Leamy LJ., Kelly SA., Nietfeldt J., Legge RM., Ma F., Hua K., et al. Host genetics and
593 diet, but not immunoglobulin A expression, converge to shape compositional features

- 594 of the gut microbiome in an advanced intercross population of mice. *Genome Biol*
595 2014;**15**(12):552. Doi: 10.1186/S13059-014-0552-6.
- 596 38. Knights D., Silverberg MS., Weersma RK., Gevers D., Dijkstra G., Huang H., et al.
597 Complex host genetics influence the microbiome in inflammatory bowel disease.
598 *Genome Med* 2014;**6**(12):1–11. Doi: 10.1186/S13073-014-0107-1/FIGURES/6.
- 599 39. Kittana H., Gomes-Neto JC., Heck K., Geis AL., Segura Muñoz RR., Cody LA., et al.
600 Commensal *Escherichia coli* Strains can promote intestinal inflammation via
601 differential interleukin-6 production. *Front Immunol* 2018;**9**(OCT). Doi:
602 10.3389/fimmu.2018.02318.
- 603 40. Rocha-Ramírez LM., Pérez-Solano RA., Castañón-Alonso SL., Moreno Guerrero SS.,
604 Ramírez Pacheco A., García Garibay M., et al. Probiotic *Lactobacillus* Strains
605 Stimulate the Inflammatory Response and Activate Human Macrophages. *J Immunol*
606 *Res* 2017;**2017**. Doi: 10.1155/2017/4607491.
- 607 41. He B., Hoang TK., Tian X., Taylor CM., Blanchard E., Luo M., et al. *Lactobacillus*
608 *reuteri* Reduces the Severity of Experimental Autoimmune Encephalomyelitis in Mice
609 by Modulating Gut Microbiota. *Front Immunol* 2019;**10**(MAR):385. Doi:
610 10.3389/FIMMU.2019.00385.
- 611 42. Paesani C., Salvucci E., Moiraghi M., Fernandez Canigia L., Pérez GT. Arabinoxylan
612 from Argentinian whole wheat flour promote the growth of *Lactobacillus reuteri* and
613 *Bifidobacterium breve*. *Lett Appl Microbiol* 2019;**68**(2):142–8. Doi:
614 10.1111/LAM.13097.
- 615 43. Singh RK., Chang HW., Yan D., Lee KM., Ucmak D., Wong K., et al. Influence of
616 diet on the gut microbiome and implications for human health. *J Transl Med* 2017 *151*
617 2017;**15**(1):1–17. Doi: 10.1186/S12967-017-1175-Y.

- 618 44. Öhman L., Lason A., Strömbeck A., Isaksson S., Hesselmar M., Simrén M., et al.
619 Fecal microbiota dynamics during disease activity and remission in newly diagnosed
620 and established ulcerative colitis. *Sci Reports* 2021 111 2021;**11**(1):1–8. Doi:
621 10.1038/s41598-021-87973-7.
- 622 45. Zhuang X., Liu C., Zhan S., Tian Z., Li N., Mao R., et al. Gut Microbiota Profile in
623 Pediatric Patients With Inflammatory Bowel Disease: A Systematic Review. *Front*
624 *Pediatr* 2021:626232. Doi: 10.3389/fped.2021.626232.
- 625 46. Pickard JM., Zeng MY., Caruso R., Núñez G. Gut Microbiota: Role in Pathogen
626 Colonization, Immune Responses and Inflammatory Disease. *Immunol Rev*
627 2017;**279**(1):70. Doi: 10.1111/IMR.12567.
- 628 47. Cao Y., Shen J., Ran ZH. Association between faecalibacterium prausnitzii reduction
629 and inflammatory bowel disease: A meta-analysis and systematic review of the
630 literature. *Gastroenterol Res Pract* 2014;**2014**. Doi: 10.1155/2014/872725.
- 631 48. Wills ES., Jonkers DMAE., Savelkoul PH., Masclee AA., Pierik MJ., Penders J. Fecal
632 microbial composition of ulcerative colitis and Crohn's disease patients in remission
633 and subsequent exacerbation. *PLoS One* 2014;**9**(3). Doi:
634 10.1371/JOURNAL.PONE.0090981.
- 635 49. Clooney AG., Eckenberger J., Laserna-Mendieta E., Sexton KA., Bernstein MT.,
636 Vagianos K., et al. Ranking microbiome variance in inflammatory bowel disease: a
637 large longitudinal intercontinental study. *Gut* 2021;**70**(3):499–510. Doi:
638 10.1136/GUTJNL-2020-321106.
- 639 50. Menezes-Garcia Z., Do Nascimento Arifa RD., Acúrcio L., Brito CB., Gouvea JO.,
640 Lima RL., et al. Colonization by Enterobacteriaceae is crucial for acute inflammatory
641 responses in murine small intestine via regulation of corticosterone production. *Gut*

- 642 *Microbes* 2020;**11**(6):1531–46. Doi: 10.1080/19490976.2020.1765946.
- 643 51. Zeng MY., Inohara N., Nuñez G. Mechanisms of inflammation-driven bacterial
- 644 dysbiosis in the gut. *Mucosal Immunol* 2017:18–26. Doi: 10.1038/mi.2016.75.
- 645 52. Tiso M., Schechter AN. Nitrate Reduction to Nitrite, Nitric Oxide and Ammonia by
- 646 Gut Bacteria under Physiological Conditions. *PLoS One* 2015;**10**(3):e0119712. Doi:
- 647 10.1371/JOURNAL.PONE.0119712.
- 648

a. **b.** **c.**

Rarefaction measure
observed species

a

b

a

Relative abundance (%)

b

Rel. abundance (median)

Healthy

Newly Diagnosed

Remission

Relapse